Machine Learning 1

Lecture 6.3 - Supervised Learning
Classification - Discriminative Models

Erik Bekkers

(Bishop 4.0, 4.1.1, 4.1.2)

Image credit: Kirillm | Getty Images

Slide credits: Rianne van den Berg

Discriminant Functions: Two Classes

- Input: $\mathbf{x} \in \mathbb{R}^D$
- Targets: $t \in \{C_1, C_2\}$ $t \in \{-1, 1\}$

Discriminant functions :

Direct mapping of input to target

- Generalized Linear Models (GLM)

non-linear activation function

$$y(\mathbf{x}, \tilde{\mathbf{w}}) = f(\underbrace{\tilde{\mathbf{w}}^T \boldsymbol{\phi}}_{\text{linear}})$$

$$\boldsymbol{\phi} = (\phi_0(\mathbf{x}), \phi_1(\mathbf{x}), ..., \phi_{M-1}(\mathbf{x}))^T$$

- Decision boundary: $y(\mathbf{x}, \tilde{\mathbf{w}}) = \text{const}_1$

$\tilde{w}^T \phi = \text{const}_2$

Discriminant Functions: Two Classes

($f(x) = x$)

- Simplest discriminant function

$$y(\mathbf{x}, \tilde{\mathbf{w}}) = \mathbf{w}^T \mathbf{x} + w_0$$

(canonical basis)

$$\phi_0(\mathbf{x}) = 1$$

$$\phi_j(\mathbf{x}) = x_j \qquad j = 1, \ldots, D$$

- Decision boundary: $y(\mathbf{x}, \mathbf{w}) = 0$

$$\underline{w}^T x + w_0 = 0$$

- Consider 2 datapoints $\mathbf{x}_A$ and $\mathbf{x}_B$ on decision boundary

$$y(\mathbf{x}_A) = y(\mathbf{x}_B) = 0$$

$$\downarrow$$

$$\underline{w}^T(\underline{x}_A - \underline{x}_B) = 0 \qquad (w \perp (x_A - x_B))$$

so w determines the orientation of the decision boundary

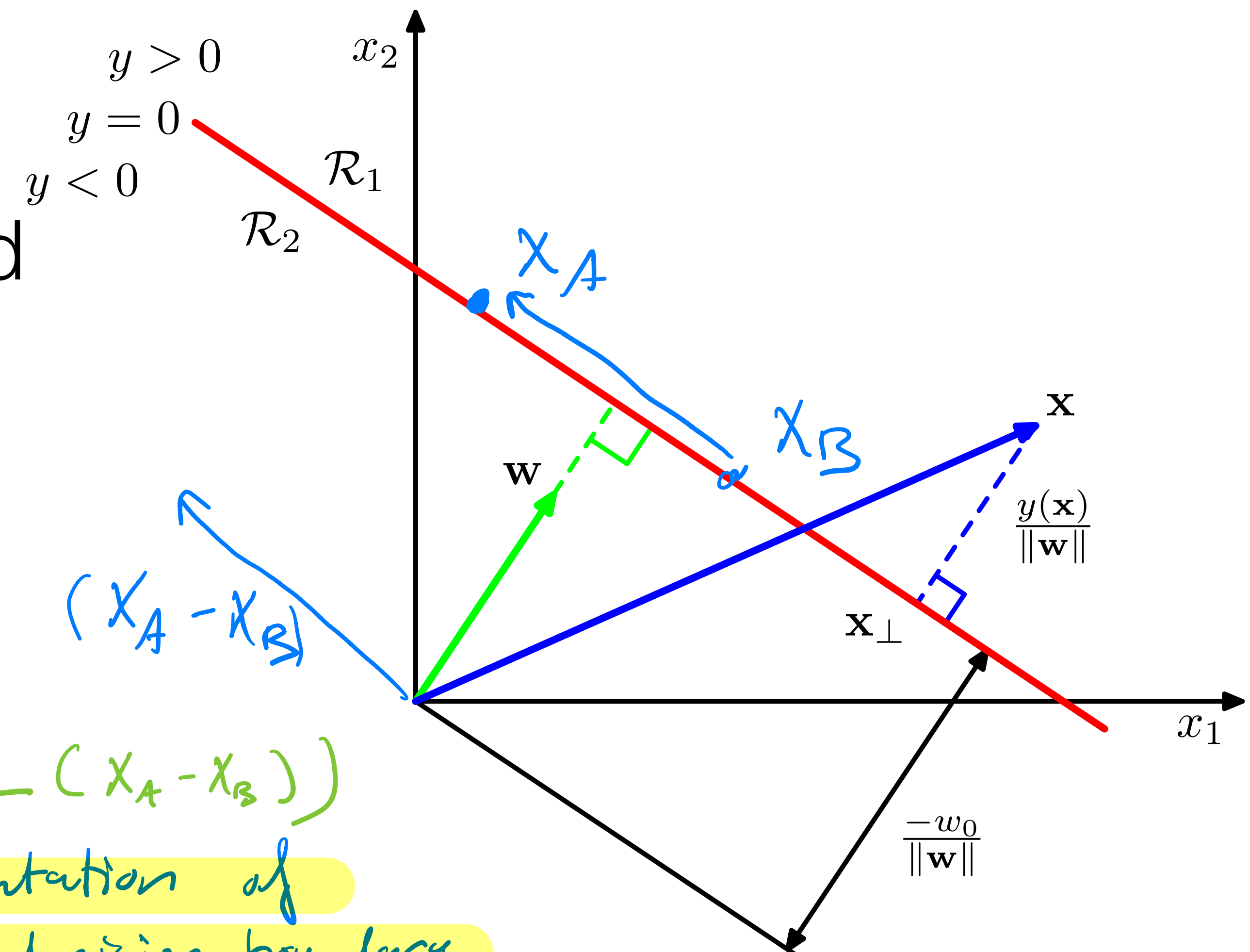

Figure: 2-class decision boundary (Bishop 4.1)

Discriminant Functions: Two Classes

- Take **x'** a point on decision surface: $y(\mathbf{x}') = 0$

$(\mathbf{w}^T \mathbf{x}' + w_0 = 0$
$\mathbf{w}^T \mathbf{x}' = -w_0)$

- Normal (signed) distance *d* from origin to decision surface

$$d = \frac{\mathbf{w}^T \mathbf{x}'}{||\mathbf{w}||} = -\frac{w_0}{||w||}$$

so w_0 shifts the boundary away from the origin!

- Normal (signed) distance *r* from general **x** to decision surface:

$$\mathbf{x} = \mathbf{x}_\perp + r\frac{\mathbf{w}}{||\mathbf{w}||}$$

$y(\mathbf{x}_\perp) = 0$

$$y(\mathbf{x}) = \mathbf{w}^T \mathbf{x} + w_0 = \cancel{\mathbf{w}^T \mathbf{x}_\perp} + r\boxed{\frac{\mathbf{w}^T \mathbf{w}}{||\mathbf{w}||}} + \cancel{w_0} = r||w||$$

$\frac{||w||^2}{||w||} = ||w||$

$\Longrightarrow r = \frac{y(\mathbf{x})}{||w||}$

so $y(\mathbf{x})$ determines the distance to the surface!

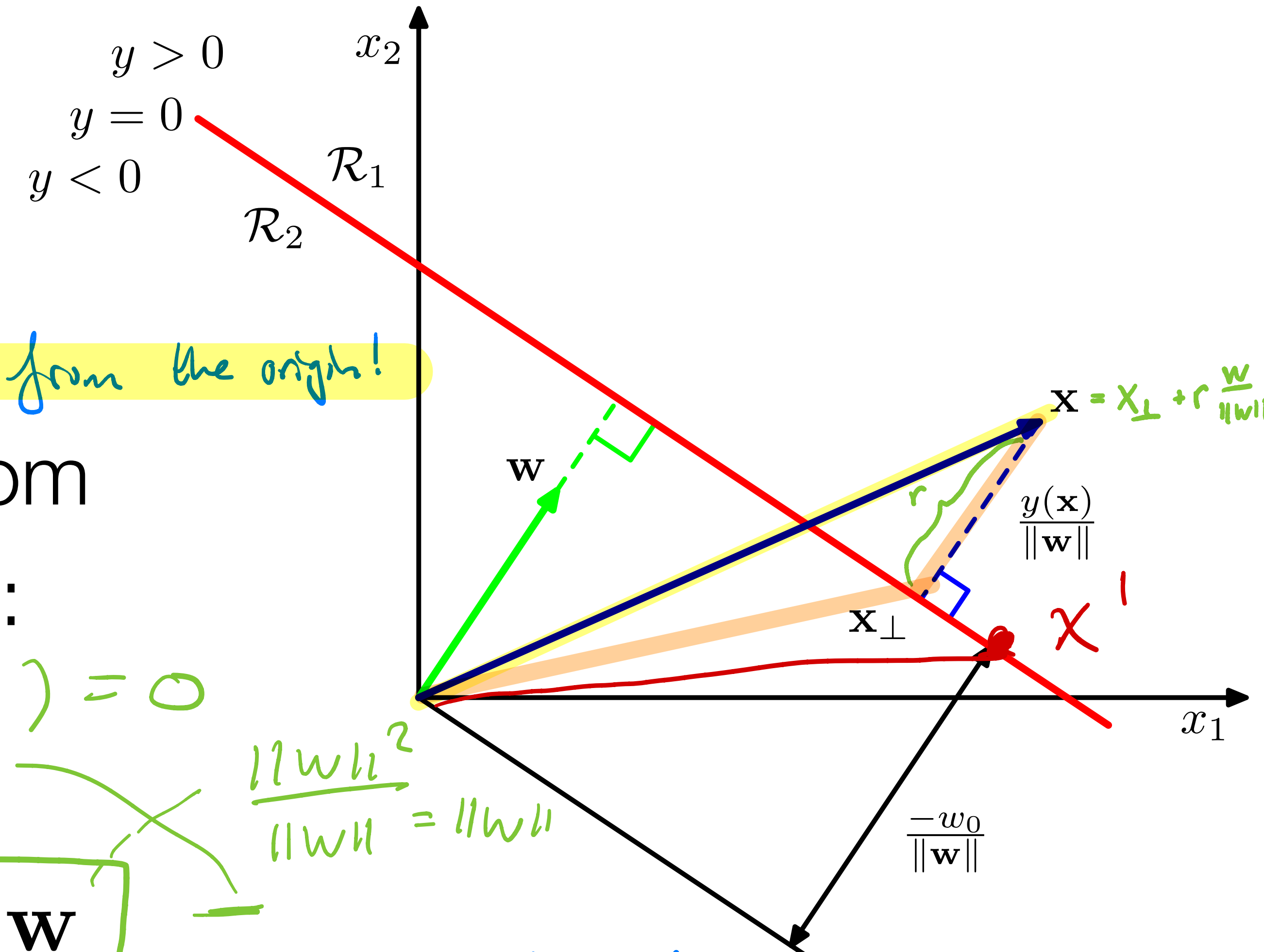

Figure: 2-class decision boundary (Bishop 4.1)

Discriminant Functions: Multiple Classes

- K-class discriminant

$$y_k(\mathbf{x}) = \mathbf{w}_k^T \mathbf{x} + w_{k0}$$

- Assign **x** to C_k if $y_k(\underline{x}) > y_j(x)$, $\forall j \neq k$

- Decision boundary between $\mathcal{R}_k$ and $\mathcal{R}_j$: $y_k(\underline{x}) = y_j(\underline{x})$

$\underline{w}_k^T \underline{x} + w_{k0} = \underline{w}_j^T \underline{x} + w_{j0}$

so linear decision boundaries

- Decision regions (for GLM) are convex